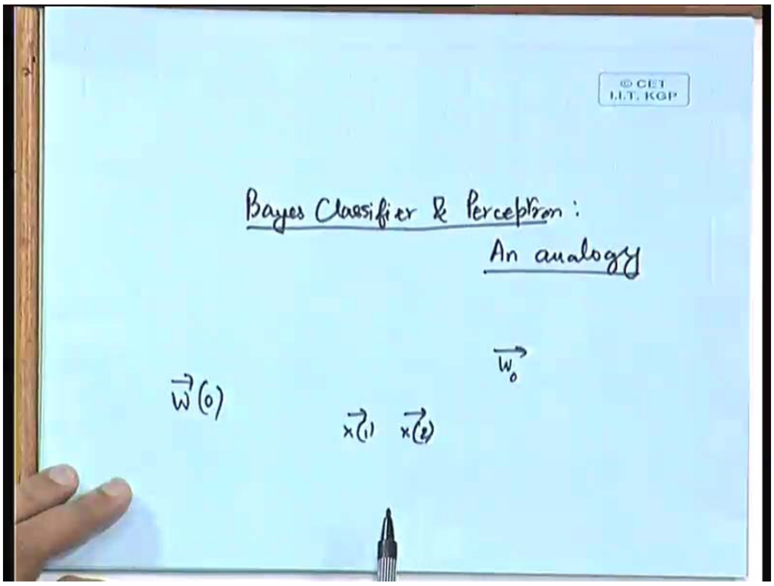
Neural Network and Applications
Prof. S. Sengupta
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 17
Bayes Classifier & Perceptron: An Analogy

Bayes classifier and Perceptron and we are trying to draw an Analogy between these two. Bayes classifier is a very well known classifier that solves the pattern classification problems in the most classical way.

(Refer Slide Time: 00:51)



Whereas, Perceptron is the elementary form of the neural network that we are considering. So, is there any relationship between these two; that is what we are going to discuss today. And before, we begin today's topic, let may just add 1 or 2 points pertaining to the last class, when we were discussing about the Perceptron convergence theorem.

So, I think one point, which I was trying to emphasize upon is that, the two alternative ways of approaching the bounds for the W 0 was that we were trying to find out first the lower bound, that must exist on the W 0. And then, we try to work out an upper bound for it and then, although both lower bound and the upper bound will be fulfilled, for the values of n, which are sufficiently small, the range would be quite small.

But, as you have seen that with the iteration n, the lower bound and the upper bound, they change and then, ultimately we have a point. Where, we are going to have some value of n max, for which the two in equations, must be satisfied with the equality sign. And so that was the n max, that we were looking for and it is that n max, within which the convergence is going to be guaranteed.

And mind you once again, that subject to some constrains and what are those constrains, number 1; the linear separability, which is of course a must, because only then, we can use Perceptron for pattern classification. Number 2; W 0 equal to 0 vector, number 3; was that eta was equal to 1. Now, let us just discuss briefly, that whether these assumptions really affect our result or not, especially the last two, the eta is equal to 1 and W 0 is equal to 0.

Now, W 0 is equal to 0 is something, that we had assumed without any a priory idea about the patterns; that means to say no priory idea about the weights. And it was logical to assume that W 0 was equal to 0. In fact, if they are not, if W 0 has got some non 0 vector value, any random initial value to start with, then the convergence will be there of course. Of course, it is not possible for you to exactly predict that whether the convergence would be faster or the convergence would be slower.

If the W 0's are selected, the first iteration W 0 if the initial weights are so chosen by shear luck. We choose W 0, such that from the initial patterns only means from X 1, X 2, etcetera, all this patterns that we are feeding. There, we have the correct classification means, by chance, if we happen to choose it, at the hyper plane of the separations that we are having between the two m dimensional spaces of C 1 and C 2.

Then of course, the convergence would be slow, that the convergence would be very fast, if it is very close to the hyper pane, planed accidentally. Whereas, if it is further away from it, then it needs more iteration, so convergence will be there. Because, of linear separability a solution of W 0 indeed exists. And that is why, the convergence will be there, whether faster or slower will be dependent upon what type of classes have we defined in our pattern set and what is the initial vector, that we choose.

Regarding eta, one point is very clear that we have chosen eta equal to 1 in this assumption. Now, supposing, we make eta smaller than 1, in that case; obviously the convergence is going to be what? Faster, slower, convergence is definitely going to be slower, so we can expect an increase or decrease of n max

Student: ((Refer Time: 06:08))

Increase of n max, because if the convergence is slower, then you are going require more number of iteration, so that means to say, that n max is going to increase. Now, again one thing is there, I think somebody was asking me, at the end of the class that, what is the guarantee, that we reach W 0. Well, I am not saying that, we have to reach W 0 only, W 0 means the optimal weight, what I suggested was that, there is a solution W 0 that exists.

Such that, they hyper plane equation W 0 transpose X is equal to 0; that gives separating solution between the pattern class C 1 and pattern class C 2. But, does not mean that this hyper plane is unique, there may be many hyper planes. In fact, it is not that W 0 is something that is unique, we can reach any solution, that is ultimately able to separate the two pattern classes will be acceptable to us.

So, there is a domain of W’s which should satisfy this relationship, so it is not that we have to reach an exact value of W 0. In fact, that is not the idea, the idea is convergence, means ultimately you come to a stage, when after feeding the patterns; you are going to have a correct classification of the patterns. Of course, one point is there, that we have analyzed it for the two pattern cases, two class cases C 1 and C 2. Of course, this may be extended to multiple classes, where we can utilize more than one Perceptron.

Supposing, you use two Perceptron, one classifying between C 1 and C 2, the other classifying between C 3 and C 4, such kind of things are possible for us to extend this whole idea of classification into more than two classes. But, the elementary number of classes, of course we are taking it to be 2. So, that was some extra discussion that I felt that would clarify any of the doubts that, you people having in your mind, regarding the Perceptron convergence theorem.

Any doubts at this stage, before we begin today’s topic that is about the Bayes classifier and Perceptron. Now, primarily we are trying to draw an Analogy, as I mention at the very beginning of when I presented this title. Now, Bayes classifier, I do not know that how many of the participants here or the viewers watching this program at a distance or having ideas about the classical pattern recognition problem. Where, you happen to learn the Bayes classifier as very basic tool for that.

In fact, Bayes classifier is one of the very popular classical approaches as for pattern classifications. And we are going to, before we draw an Analogy about Perceptron, because Perceptron we already know. But, just a brief over view of Bayes classifier, I am giving, in case most of the Perceptron do not have ideas about the Bayes classifier. In case, you do not have, I think some amount of over view would be necessary.

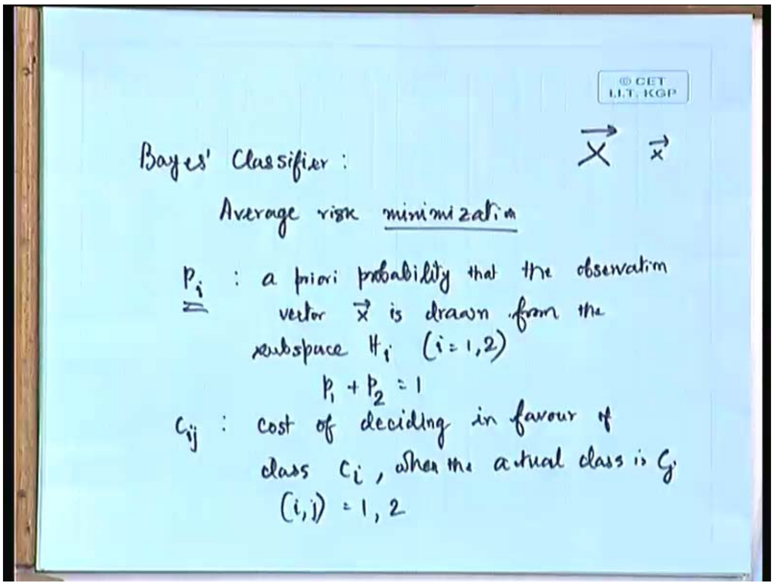
And here also, we are going to consider the simplest case of two pattern classifiers that is the classes belonging to this are C 1 and C 2. Now, one thing is there, that for the case of Bayes classifier, it is shown, that when the patterns are having a Gaussian distribution. When, the patterns are chosen from a Gaussian distribution in such cases, it is shown that Bayes classifier becomes a linear classifier problem.

Bayes classifier, in the log likelihood sense, which we are going to come later is going to be linear classifications solution, exactly like the way our Perceptron dose. Of course, the difference is that, in the case of Perceptron, we are not making any a priory idea about the probability density functions of the patterns, belonging to the class C 1 and C 2. Those distribution could be anything arbitrary, the only thing that we are stipulating for that is, that the pattern classes are linearly separable.

Whereas, in this case mind you, that when I am talking of the Bayes classifier, then the pattern classes, that we are having. Since, they are Gaussian distributed and the both the C 1 and the C 2 classes will have the patterns distributed in a Gaussian way, in the manner of a Gaussian probability density function. They should necessarily be over lapping, because theoretically two Gaussian functions, we cannot have as non overlapping case.

So, they lie one of the difference, but let us have the basic introduction to the Bayes classifier. Again, as I told you just now, that it is based on the two classifier stages.

(Refer Slide Time: 12:15)



Now, here the basic thing that we are going to do for Bayes classifier is that, it is based on what is called as the average risk minimization. Now, I tell you what it means, we are going to define a risk function that considers weighted sum of correct classification and the misclassification. Of course, when we make a weighted sum like that, in that case, more weights should be given to the weights pertaining to the misclassification.

Because, when I say minimization; that means to say that, we have to minimize that risk functional. So, minimize that risk functional; obviously means that, that should have an effect in minimizing the misclassifications that we are getting out of the system, out of the classifier system. So, definitely, when we are putting up a weighted sum, the idea should be is that more weights should be put to the misclassification.

So, just to present that simple model of the risk function, that we are going to define, we have to define beforehand, just a few elementary quantities. First is a probability Pi, that we are saying, which is an a priory probability, that an observation X vector. In fact, what the X vector that we drawing is out of a random vector capital X. So, we take a random vector capital X.

So, this a random variable in a vector form and small x, that we are considering is just an instance of that random vector. So, small x mind you is an instance of the random vector x. So, what we are defining probability Pi, it is some kind of an priory probability, whose idea you might have or you may not have the a priory idea. In fact, this probability, the initial assumption of probability is not that something, which is going to matter in a great way later on, we are going to hop upon that point later.

So, it is an a priory probability, that the observation vector x, observation vector is small x, because that is an instance of capital X is drawn from the subspace H i. Where, I is equal to just 1 and 2, because we are having two classification problems, we are only having two classes, so H i's, i can be either 1 or 2. So, what does P i says that, it just gives us a probability, that it is drawn from the class H I. That means to say, it is drawn from H 1, the probability is P 1 and if it is drawn from H 2, the probability is P 2.

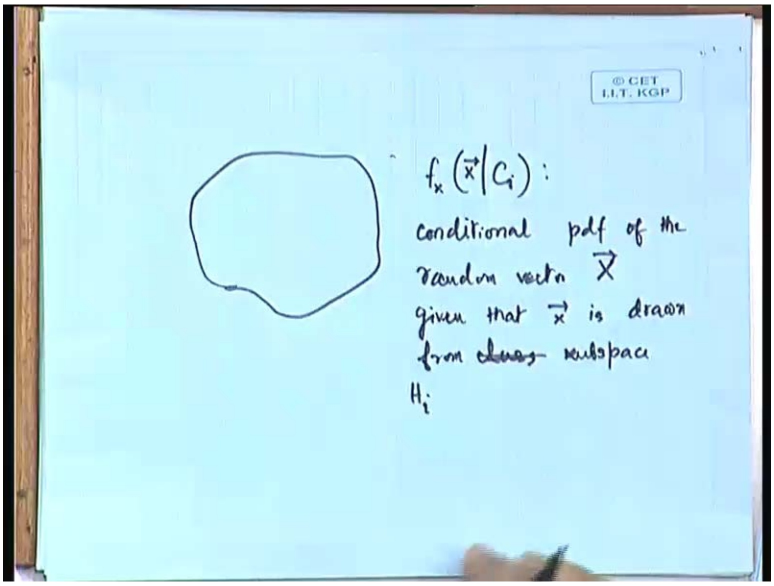
In absence of any a priory idea about the probabilities, lot of times, what people will do is that, take P 1 and P 2 both equal to half. In fact, that also quite a reasonable assumption, but I am not going into the specific case of P 1 and P 2 both equal to half, let say P I, that the probability that is drawn from this subspace H i. Now; obviously, here by the very basic nature of this definition, because it is a two classifier problem, P 1 plus P 2 should necessarily be equal to 1, so that is one of the definitions, that we are having.

And the next is C i j, which is a cost function, in fact C i j is the cost of deciding, means cost of classifying, deciding in favor of class C i. Class C i. When the actual class is C j and here i and j both could be 1 and 2, meaning what, what is C, so that means to say that, we can have four such cost functions C 11, C 12, C 21 and C 22. What is C 11, C 11 means that, it is actual class is 1 and the classification that we have made is also 1, which is a correct classification.

What is C 22, actual class 2, the classification made by the classifier is 2, so it C 22 is also a correct classification, correct classification into class 2. But, what does C 12, C 21 indicate, they indicate misclassification. So, C 12 means that, it is drawn from the class C 2, but we have assigned it to the class C 1. And like wise, C 21 will mean that, it is drawn, the pattern is drawn from the class C 2. But, actually it will, it is classified into C 1, again C 1 it is drawn and it is classified into C 2, so that shows a misclassification.

So, C 21 and C 12, they are misclassification cost functions, where as C 11, and C 22, they are correct classification cost functions. So, we should necessarily have C 11 and C 22, as very small quantities, whereas, C 12, and C 21 would be having relatively larger values.

(Refer Slide Time: 18:53)



Now, other than a priory probability and the cost functions, we are also going to define the probability density function, because just remember one thing, that supposing, here this is the class C i, that we have got. And we have taken just a subspace out of the class C i which we are calling as the H i. Now, remind you, the subspace or the entire space C i, whatever I consider that have to be drawn out of an m dimensional space.

So, naturally when I draw a space like this, imagine it is an m dimensional space that we are considering. Now, in that m dimensional space, all the random variable X vector that we are considering will be distributed within this m dimensional space. So, there will be some form of a distribution. So, if it is having a Gaussian distribution, then what can you expect, that most of the instances of the patterns, they will be clustered more around the mean.

There will be some mean vector, that will be existing obviously and the clustering will be more towards the mean, there will be more number of patterns that could be found from that mean area. And then, as we go away from the mean, it will fall off in a Gaussian way and remember, that in this case, the Gaussian that I am trying to define is an m dimensional Gaussian.

So, whatever ideas we are having about one dimensional Gaussian function that is what we are considering most of the times in our statistic analysis or people have extended it to two dimensions. Those people, who are working with two dimensional problems like image processing, but in this case, again is it is a multidimensional Gaussian, that we have to consider.

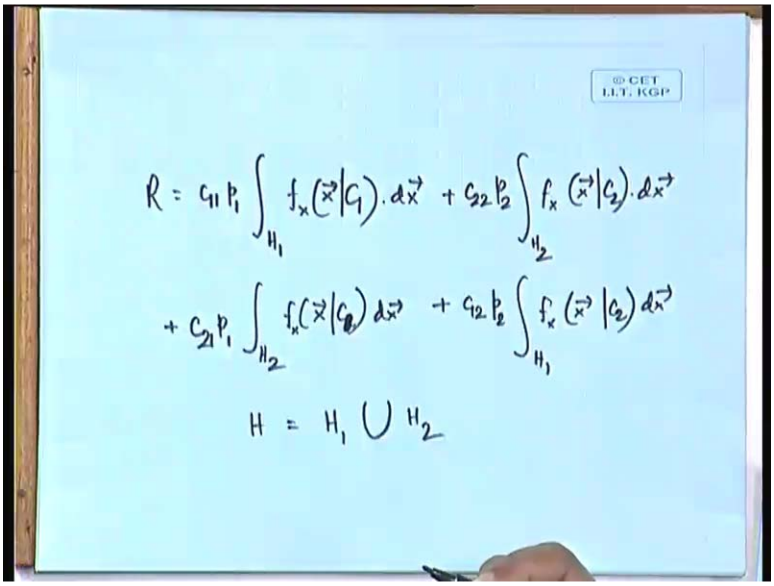
So, associated with that multidimensional probability density function, right now we are not making any a priory assumption, that it is indeed Gaussian. In fact, Bayes classifier in a general way, we can consider any probability density function. But, I was trying to tell you at the beginning itself is that, when the probability density function happens to be Gaussian, then the Bayes classify classifier becomes a linear classifier. But, in general we are going to consider a probability density function, which we are going to call as f x.

And in fact, we are going to define the probability density function f x of x in a conditional way, that is this x, f x, x given C i, this means to what, that this is a conditional probability density function, so this is a conditional pdf of the random vector x. So, this is the random vector capital X, given what, given that the instance small x is drawn from class C i, drawn from or we say better from subspace H i. So, this is the conditional pdf.

So, after defining these three quantities, that is the probability, the cost function and the conditional probability density function, now we go over to the definition of the risk functional. So, the risk functional, in fact will be a combination of how many terms, can you guess, how many terms should be there in the risk functional, four terms will be there, one is involving C 11, one is involving C 22, they are the correct ones. And C 12 and C 21, which are the misclassification costs with which we have to associate.

So, definitely what we have to do is that now, that we have idea about the conditional probability density function, we will have to ultimately define the integration of this f x with the d x vector. So, this f x again that we are defining is mind you in the m dimensional space, so it is the m dimensional conditional pdf; that we are considering.

(Refer Slide Time: 23:40)



So, the risk functional R, we can say will be defined as follows, will be equal to C 11, times P 1, P 1 means it is the probability, that the observation vector is drawn from the class H 1. So, it is C 11, P 1 an integral of f x, x vector being drawn from the class C 1, we are drawing the X vector from the class C 1. And then, this multiplied by d x, dot d x, where d x is nothing but the elementary vector x, that we are going to consider.

We saw this is drawn actually from the space C 1 and where have we classified, we have also classified it into C 1. So, because we have classified into C 1, which is the subspace over which, we should integrate this function H 1 or H 2. Yes, majority is saying H 1, but somebody said H 2. So, an analysis what it that, why did you say, it is H 2, it is not H 2, in fact.

In this case, there is no question of H 2, because we are drawing it from C 1 putting also into C 1 a correct classification. So, this pdf that we are considering should definitely be defined under the H 1, so we have to integrate this probability density function over H 1. So, that is one, that is the first term that we are getting, the second term that we are getting, again before writing the misclassification terms, I am writing the correct classification once.

So, here we consider C 22, which is the cost associated with the classification with 2, drawing from 2, 2 and classifying into 2. So, it is likewise C 22, P 2 integral f x, x given C 2 d x and in this case, the integration is going to half over H 2, that is correct.. So, about this two terms, we do not have any problem, but we are also going to have two extra terms. In fact, the contribution of those terms will be more in the risk function, because as I said that, we are going to have C 12, C 2 once value to be higher.

So, what we are going to consider is, that it should be C 21, P 1 and why P 1, because we have drawn it from the class 1. So, that is why P 1 will be associated it and we have drawn from class 1 and put into class 2, misclassification. So, here what should be this f x, x given correct, x given C 2 d x. And where are we going to put, no, not x given C 2, sorry, we are drawing it from class 1, only.

So, it is x given C 1, so again I was guided by wrong feedback that was coming from the audience. So, it is x given C 1, it is drawn from 1 and put into 2. And what is going to be the integration subspace here, sure. So, it is a divided house, half of the people saying H 1, half of the people saying H 2. Well, you have to integrate over the space, where you are putting into the classification. So, this should be integrated over the space H 2 not H 1.

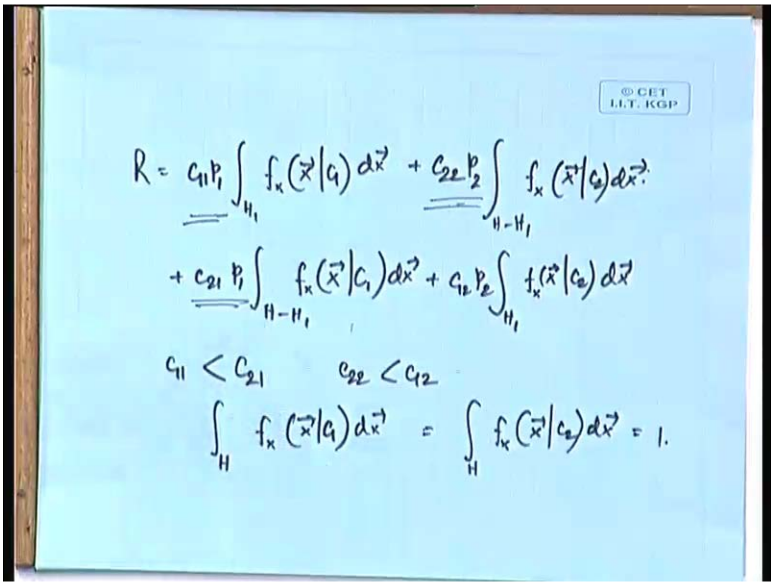
And the other term, the 4th term that we have to consider is C 12, now I think, even if you do not understand the thing, you can write the 4th term by looking at the simple Analogy of it. Because, we have put C 21 here, we have to put C 12, here, because we have put P 1 here, we have to put P 2 here, because we have put the H 2 here, we have to put integrating over H 1 here and x given C 1 here, so there we have to give x given C 2.

Even, if we understands, do not understands we can write it down, anyway the logic remains the same, so it is C 12, P 2 integration over H 1 f x, x given C 2 d x. So, these are the four terms, so now what is the total observation space that we are having. We are having two spaces, one is H 1, the other is H 2 and the total observation space that we are having is the H 1 union of H 2, which we are going to call as the set H.

Now, once we define a set H, as a union of H 1 and H 2 definitely, it is possible for us to represent let say H 2 could be represented as H minus H 1. H 1 could be represented as H minus H 2; that is very clear from this definition itself. But, when I consider a union set, now if I draw the instance of the X vector and if I asked that, whether the probability that is drawn from the subspace H, that is of course, should be equal to 1.

And if we consider integral f x over the H space, over the entire H space, that integration is going to result in 1, integral f x, x, d x, if we consider that going to be going to 1. So, accordingly, what we can write down is that, this risk equation that we have written down can now be modified as follows.

(Refer Slide Time: 30:14)



So, this the risk equation can be just simply rewritten as C 11, P 1, the first term, I am writing. So, I am going to write all this four terms and you please verify that, what I am writing is very much consistence. So, it is C 11, P 1 integration over H 1, f x, x given C 1 d x plus C 22, P 2 integration of f x, x C 2 d x and all that I did was that instead of H 2, I am going to write simply H minus H 1.

And likewise, the next term is going to be C 21, P 1 integral instead of H 2, I write H minus H 1, f x, x given C 1 d x plus C 12, P 2 integral H 1 f x, x given C 2 d x. So, just I did nothing extra, I only replace this H 2's by H minus H 1. But, having done that, having defined H 2 in that manner; that means to say that when I consider these terms integration over H minus H 1; that could be regarded as if to say, that we integrated over H and the integrant that we are getting, from that we can subtract the integrant that will be defined over H 1.

So, we can do that and if we do that, then integral over H, of course integral over H f x d x is going to be equal to 1. So, there will be a term that comes with the minus sign, which will be associated with the H 1 space. So, now that we have got, here H 1, here also another term with H 1 with a negative sign. Here again, H 1 with a negative sign and here H 1 with a positive sign.

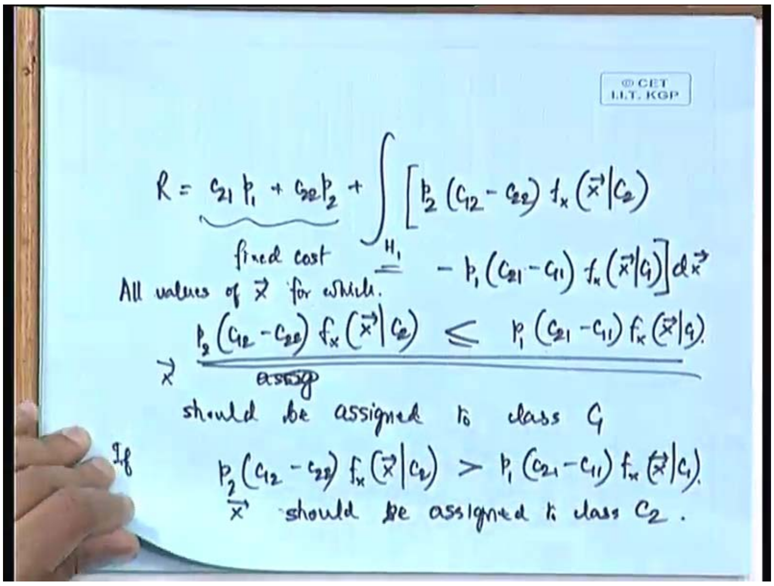
If I combine these four terms, all the integrations being defined over the space of H, so definitely we are making use. So, we are going to make that simplification and also noting very much that C 11 is less than C 21. Because, what is C 21, C 21 is that, it is drawn from the class 1 put it into class 2 and C 11 is drawn from class 1 and putting into class 1. So, it is correct classification of 1 and this is incorrect classification of 1.

So, C 11 is obviously less than C 21 and C 22 is less than C 12; that is one of the restrictions that we are having. And another is that, because H is the total space, so that is why integral f x, x given C 1, d x, whether we call x given C 1, d x or whether we call x given C 2, there this integration is going to be equal to 1, as you very rightly said. So, by applying these, what we are going to get is as follows.

You see, just try to identify some of the term very quickly, one is C 11, P 1 and then, it is integral of H 1 that we are getting and here it is C 22, P 2 and then, we are getting integral H minus H 1. So, now; obviously, integral C 22, P 2 that you are getting as integration of H; obviously means that, for that term, we are going to have C 22, P 2 as the term that comes out directly. Because, integration within that, is equal to unity integral of H, when we are considering.

So, there is a term, which we are getting directly as C 22, P 2 and there is another term, that we are getting, which is C 21, P 1. So, C 22, P 2 and C 21, P 1 will be two terms, which will emerge directly. And the rest of the terms and the rest of the integration will be done over what space, H 1 space.

(Refer Slide Time: 35:04)



And if we write these separately, what we get is that, the two terms that comes out without integration or where the integration becomes trivial, equal to unity, those two terms are C 21, P 1 plus C 22, P 2 as we discussed just now. And there will be an overall integration, under the H 1 space, where we are going to have, let me write down and then you can quickly verify. That nothing wrong is being written P 2 into C 12 minus C 22, f x, x given C 2 minus P 1, C 21 minus C 11, f x, x given C 1 d x.

Now, how this term comes is pretty clear, you see here, that why I am getting this P 12, C 2 term. The P 2, C 12 term is coming actually from here, P 2, C 12, because we are considering all the integrations within H 1 space, integrations over H 1. So, there we definitely get a C 12, P 2 term with a positive sign and as a negative sign, we are going to find what, C 22, P 2.

So, if we take P 2 common, it is C 12, minus C 22 and in both this cases, this one as well as this one, it is x given C 2. So, we get this term, no objection and then another term that we are getting is that we get C 11, P 1, for this H 1 term, we are getting C 11, P 1. And the other, that we are getting is minus of C 21, P 1 and both this will be having a common thing as f x, x given C 1, so what we got is quite correct.

So, now the first two terms here that we are getting these two are fixed costs, fixed costs in the sense that, they are not dependent upon the probability density function. Whereas, the quantity that we have written, under the integrant sign, under the integration sign is very much a probability density function dependent. You can see that, here the f x, x given C 2 and f x, x given C 1 both the terms coming into picture Whereas, this is a fixed cost, now what is our objective, our objective is to minimize the cost.

So, we have got a fixed cost component, already built into it, why it is fixed cost, because we have decided upon some priory probability P 1 and P 2. And we have also decided about some cost function, no matter, how we arrive at the cost function, but C 21 and C 22, we have chosen, so it is definitely fixed. Now; that means to say that our objective would be just to consider a minimization of this expression.

Now, look at this one, P 2 of this minus P 1 of this, if this term is going to be less than 0, if this whole quantity that we have written, that means to say, that if P 2 into C 12 minus C 22, f x, x given C 2 is less than or equal to let us say P 1, C 21 minus C 11 into f x, x given C 1, if that is the case means 0 or negative. And if you do not take the equality sign, you can say that it is the negative thing.

But, if that is negative, then where are you going to put it into, you are definitely going to classify it into the class C 1, why because this integration, please note, that this integration is defined over H 1. And we would very much invite a negative term into it, because if this term becomes negative, then you are reducing the cost function. So, definitely if this is a thing, that is contributing to the H 1 space; that means to say, that if this is classifying into H 1.

Then, if this negative relationship is satisfied, if this quantity is less than this quantity, then the classifier should put it into the class C 1. So, if this is true, so we can say, that all values of X vector for which this quantity is less than this, those access should be assigned. So, for all values of x, for which this is true, x should be assigned to class C 1, you are absolutely correct.

And of course, the opposite is also going to be true; that means to say, that if we happen to find that P 12, C 12, minus C 2 remains. If x is drawn in such a way that P 12, C 12, minus C 22, f x, x given C 2 is greater that P 1, C 21 minus C 11, f x, x given C 1. If this had happened, then we would have got a positive quantity and if we get a positive quantity into this.

We should not be putting into H 1, why should we, because if we put it into H 1, then it increases the risk cost. So, we should not put it into H 1, we should not increase the risk that way. So, in that case and because, we have got only two choices, we can either classify it into C 1 or into C 2. So, if this is true, that is this is greater than this, it should be, then x should be assigned to C 2.

In fact, one can say that, although I said that, this one is assigned to class C 1, even if we take this with a negative sign totally. If I say, because one thing is there, that if it is having an equal to sign, then it does not matter, that whether you put it into the class C 1 or class C 2, because the integrant in this case is going to be equal to 0. So, it is not going to add up to or reduce the risk function cost, so this is the classification criteria.

That is this P 2 quantity less than this P 1 into this quantity, put it into the class C 1 or otherwise you put it into the class C 2. Now, this is quite a manageable form that we have got. Now, what we are simply going to do is, that we are going to just define it in a bit of other way. Let us take the less than sign only, forgetting about the equality, because equality you can assign to any class, it does not matter, yes please

Student: ((Refer Time: 43:45))

If both are equal, yes

Student: ((Refer Time: 43:48))

Integration will be 0

Student: ((Refer Time: 43:52))

That is correct,

Student: ((Refer Time: 43:56))

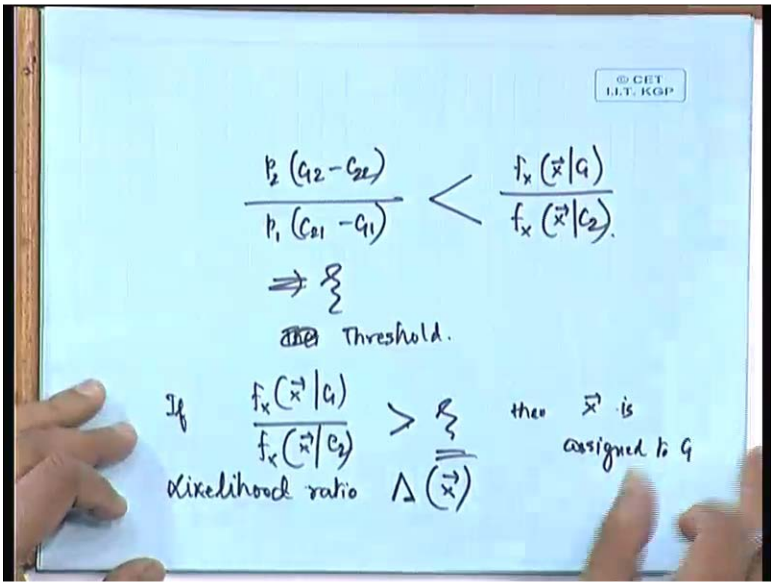
Cost C 21 is greater than C 2 C 2 as here

Student: ((Refer Time: 44:04))

No, you see, there is there is no question in risk increase, because here, I am drawing the see this risk is not dependent upon the drawing of the function x vector. We are picking up the function x vector. Now, the point is that because we within the integration sign, we have got distribution of this x vector, so that is why it is affecting. Whereas, here whether you draw x vector from C 1 or whether you draw X vector from C 2, that does not matter here, the pdf is not coming into. So, that is why, we are saying that, it is a fixed cost.

No matter whether you classify it into C 1 or you classify into C 2; that is not going to affect cost, so in that sense, it is fixed cost, is it understood. So, now, this expression the left hand side and the right hand side is lightly rearranged.

(Refer Slide Time: 44:59)



So, that I can say that P 2 into C 12, minus C 22, by P 1 into C 21 minus C 11, you see what I have done is, that I have taken on the left hand side and brought this into right hand side, so I have to divided this P 2 into this quantity, by P 1 into this quantity. So, that at the other end of the lesser sign on the right hand side of this in equation, we are going have f x, x given C 1 by f x, x given C 2, this is a very interesting form.

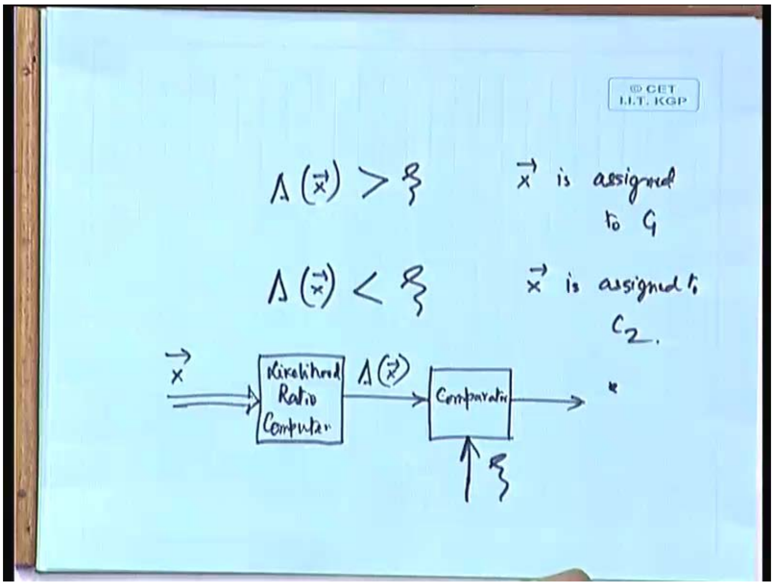
Now, here you see, what is the first term is; first term is again something, which is fixed in the sense, that we have assigned a priory probability P 1 and P 2. And we have also assigned the cost function C 12 and this things. So, this is definitely a fixed thing and another thing, you note that C 1 to minus C 22, what will be the sign of this, positive. Because, C 22 is having a lower post, C 22 means, it is drawn from 2, put into 2, C 12 means it is drawn from 2 and put into 1.

So; obviously C 12, the cost function for this classification is going to be higher than the cost function of correct classification. So, C 12 is greater than C 22 and likewise C 21 is greater than C 11. So; that means to say that this quantity that we are having is going to be positive or negative, definitely positive. So, this is called as xi, so this is defined as quantity called xi, which is called as the threshold.

And very clearly, it is the threshold, in the sense that, if we can say, that if this quantity that is f x, x given C 1 by f x, x given C 2 is greater than xi, then what classification, then x vector is assigned to, then x is assigned C 1. So, definitely xi indeed acts as a threshold of it. If it is less than, we are going assign it to C 2, if it is equal to assign it to any class does not matter.

Now, this quantity is a ratio of what, ratio the pdf; that means to say that, this pdf is what, likelihood of C 1 and this is likelihood of C 2. So, this we can call as a likelihood ratio, so this is what is called as the likelihood ratio. And this is defined in terms of the likelihood or what likelihood of X vector. So, that is why, we should write it as the likelihood ratio, we are expressing as the capital lambda of x vector. So, what we are going to have a mind you even this also is going to be positive, correct.

(Refer Slide Time: 48:35)



So, likelihood ratio is positive and this lambda x, if that is greater than the threshold xi, then x is assigned to C 1. And if lambda x is less than xi, then x is assigned to C 2 equal to you assign to anything does not matter. So, this is the essence of the Bayes classifier problem. And we can simply, just assign, show it represent in the form of a very simple block diagrams.

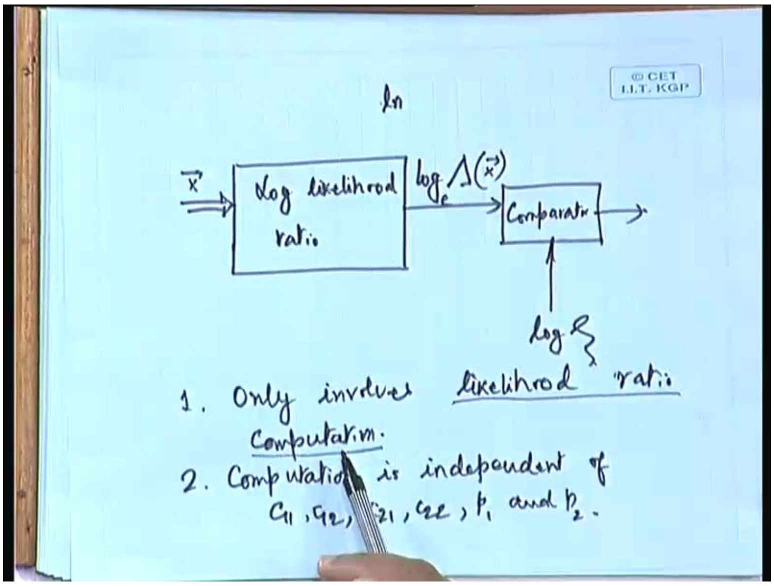
So, in that, what we are going to have is that, we are going to have a block, the input to which is the x vector and in this block, we are going to calculate the likelihood ratio. So, here we are going to compute the likelihood ratio. So, we can say likelihood ratio, computer. So, in this case, what are you going to get output of this block is nothing but this lambda x.

And here, there is a decision box; that will give the classification output, so here the output will be either C 1 or C 2. So, in this case, we are simply going to have, what is called as a comparator. So, there will be a comparator and it will compare this lambda x with what, with xi. So, lambda x will be compared with xi and in that case, we will be assigning x vector to the class C 1. So, it will be assign to class C 1, if lambda x is greater than xi or otherwise it will be assign to class C 2.

Now, this is the simple way of computing the likelihood ratio, simply looking at the ratio of the probability density function. But, one thing is clear, that we often need to compute, we want to utilize the dynamic range of this likelihood ratio. So, that is why, rather than computing the likelihood ratio directly, it is very often convenient to take the logarithm of the likelihood ratio and then make a comparison based on the logarithms.

So, what we are going to do is that, we are going to compute the natural logarithm of lambda x, as well as natural logarithm of xi, I am going to compare these two. And in such kind of cases, the likelihood ratio that we are going to compute will be refer to as the log likelihood ratio, so just putting it in a logarithmic way.

(Refer Slide Time: 51:58)



We can say that your x vector is the input and then you are going to have here a log likelihood ratio, computer. And then, you are going to get here log of lambda x, log mind you is the natural logarithm, i could write it as l n, does not matter. So, here always understand that, it is a natural logarithm and here this is the comparator and here we are going to have the log of this xi function and here will be decision.

That if log of lambda x is greater than log of xi, in that case assign, it to the class C 1, otherwise you assign it to the class C 2. So, two very interesting things, that we are finding is that, one is that the data possessing of these classifier, only involves likelihood computation, because other than that, there is no other computation. So, it only involves the likelihood ratio computation.

Now, here the likelihood ratio that we are computing, that is completely independent of whatever C 1, C 2, etcetera, we were C 11, C 12, C 21 C 22, etcetera, we are chosen, it is completely independent of that. All that we need is the probability density function of these two. Pdf of f x given is C 1 and pdf of x given C 2, other than that, we do not required those.

So, computation is independent of C 11, C 12, C 22, C 21, C 22, P 1 and P 2, all the a priori assumptions that we are making. But, that says only that this likelihood ratio, a log likelihood ratio is not dependent upon this factors, but definitely the decision is independent upon C 11, C 12, C 21 and C 22. Why, because ultimately these parameters indeed affect xi.

So, if it affects the threshold, then naturally the decision making is very much influenced by the choice of these parameters, but the computational part is only this. So, this is great advantage of the Bayes classifier. Now, in this class, actually I intended to show that when the probability density function f x of x is going to be Gaussian. Then, we could show that the classification decision that we are having; that means to say computing log of lambda x and computing log of xi and making a decision based on that.

The decision boundary equation that we are going to get is a linear equation that we are getting, when, probability density function is happens to be Gaussian. I intended to do that in today's lecture, but I have to shift it to the next one. Because, you know always one thing happens, that which at least you the distant viewers will not be able to see that. I am afraid of one thing, when I come at this lecture, that the person always comes to me with sign like this.

So, that means to say, that I have got only 5 minutes and I think the person is already alerted me 5 minutes back, that my time is going to be over. So, I should not be eating away into the time and pause the difficulties. So, we are going to postpone this decision; that means to say, they further analysis of the Bayes classifier in a assuming the Gaussian probability density function, showing it to be linear in the coming class.

Thank you very much.